3-2009

# Testing Conditional Uncorrelatedness

Liangjun SU
*Singapore Management University*, ljsu@smu.edu.sg

Aman ULLAH
*University of California - San Diego*

**DOI:** https://doi.org/10.1198/jbes.2009.0002

# Testing Conditional Uncorrelatedness

**Liangjun Su**
Guanghua School of Management, Peking University, Beijing, China, 100871 (*lsu@gsm.pku.edu.cn*), and School of Economics, Singapore Management University, 90 Stamford Road, Singapore, 178903 (*ljsu@smu.edu.sg*)

**Aman Ullah**
Department of Economics, University of California, Riverside, CA 92521-0427 (*aman.ullah@ucr.edu*)

We propose a nonparametric test for conditional uncorrelatedness in multiple-equation models such as seemingly unrelated regressions (SURs), multivariate volatility models, and vector autoregressions (VARs). Under the null hypothesis of conditional uncorrelatedness, the test statistic converges to the standard normal distribution asymptotically. We also study the local power property of the test. Simulation shows that the test behaves quite well in finite samples.

KEY WORDS: Conditional heteroscedasticity; Local polynomial estimator; Nonparametric multivariate regression; Seemingly unrelated regressions; Vector autoregressions.

## 1. INTRODUCTION

There is extensive literature on econometric models and their estimation with unconditional correlations, for example, serial correlation in time series regression models (Hamilton 1994), spatial correlation in cross-sectional models (Anselin 1988), and correlation in the system of multiple equations models, which include seemingly unrelated regressions (SURs) by Zellner (1962), vector autoregressions (VAR), multivariate volatility models, and heterogeneous panel data models (Pesaran 2006). Smith and Kohn (2000) also considered the extension of Zellner's SUR model to the nonparametric regression with unconditional correlations. There are also some existing parametric tests to verify if the correlations (dependence under normality) are statistically insignificant. This testing is useful since under zero correlation it is generally easier to estimate econometric models as compared with when we estimate them under the presence of correlation. The well-known parametric tests are the Durbin-Watson and the Box-Pierce tests for serial correlations, the Moran (1948) test for spatial correlation, and the tests for correlations in the system of multiple equations by Breusch and Pagan (1980), Kariya (1981), Frees (1995), Dufour and Khalaf (2002), Pesaran (2004), Ng (2006), and Pesaran, Ullah, and Yamagata (2006). However, all these estimation and testing procedures are for models with unconditional correlations.

In recent years, several parametric and nonparametric models have been suggested with conditional correlations. While the unconditional correlations imply constant correlations over the observations, the conditional correlations vary over the observations with respect to variables in the conditioning set. Also, an unconditional uncorrelated process may still be a conditionally correlated process (Anatolyev 2002). In many models, the conditional correlations arise because of conditional covariances or conditional heteroscedasticity. For example, the conditional correlations have been studied in the linear parametric SUR models due to conditional heteroscedasticity (Duncan 1983; Singh and Ullah 1974; Srivastava and Giles 1987; Mandy and Martins-Filho 1993). These authors also provided the estimation of such models when the conditional heteroscedasticity is parametrically specified. If the heteroscedasticity is nonparametric, then the extended method of Robinson (1987) can be used. Recently, Welsh and Yee (2006) considered nonparametric SUR models with the conditional correlation of errors. We note that when we have the conditional uncorrelatedness or weak correlatedness, SUR can be analyzed by estimating each equation separately. Conditional uncorrelatedness, under nonnormality, is weaker than conditional independence, which was recently reexamined by Su and White (2007). Thus, the test for conditional uncorrelatedness is useful.

The argument about SURE also extends to other multiple equation models such as simultaneous equation models and parametric and nonparametric VAR models (e.g., Härdle et al. 1998). Here, we can also test for conditional uncorrelatedness before we estimate the equations jointly in a VAR system. Another important development has taken place in modeling conditional correlation/covariance in financial econometrics literature (Engle and Kroner 1995; Alexander 2001; Engle 2002; Engle and Sheppard 2002). These authors provide various multivariate GARCH models for the conditional correlations across the financial markets. On the other hand, Fan et al. (2008) assumed that the multivariate financial time series is a linear combination of a set of conditionally uncorrelated components, which overcomes several drawbacks of early models. All these studies provide a reason for the testing of conditional uncorrelatedness, especially since the acceptance of conditional uncorrelatedness leads to analyzing the multivariate GARCH model simply by estimating the univariate GARCH models.

Motivated by the preceding considerations, the purpose of this article is to propose a test for conditional uncorrelatedness. To the best of our knowledge, our test is the first nonparametric test for conditional uncorrelatedness in the literature. There are several key features that are associated with our test. First, our test is nonparametric. It is well known that parametric tests may be powerful for certain types of alternatives and perform well

**Published in Journal of Business & Economic Statistics, 2009, Volume 27, Issue 1 pp.18 - 29**
**https://doi.org/10.1198/jbes.2009.0002**

in the case of correct specification; they can lead to misleading conclusions in the case of misspecification. In contrast, smoothing-based nonparametric tests work well for a wide class of alternatives and yield good power in a variety of circumstances despite the fact that they are subject to the curse of dimensionality. Second, our tests allow for weak dependence in the data. This is important when one applies our test to time series data, say, in a VAR system or in multivariate volatility models. Third, even though we focus on testing for conditional uncorrelatedness in the nonparametric regression framework, it is directly applicable to the residuals obtained from a parametric regression under weak conditions. Fourth, as we argue in Section 3.4, our proposed conditional test is analogous to the normal approximation version of the LM (NLM) test statistic for unconditional correlation. The differences are that the NLM test was developed for the parametric regression models, and it is based on the estimate of the unconditional correlation, whereas we consider the nonparametric regression model and our test is based on the estimate of conditional covariance.

The rest of the article is organized as follows. We introduce our testing framework in Section 2 and study the asymptotic properties of our test statistic in Section 3. In Section 4, we provide a small set of Monte Carlo experiments to evaluate the finite sample performance of our test. Section 5 includes conclusions. All proofs are relegated to the appendix.

## 2. BASIC FRAMEWORK

In this section, we first state the hypotheses and then introduce the test statistic.

### 2.1 The Hypotheses

We consider the system of $P$ nonparametric regression equations:

$$Y_{ji} = m_j\left(X_{ji}\right) + U_{ji}, \quad j = 1, \ldots, P, \quad i = 1, \ldots, n, \tag{2.1}$$

where $n$ is the number of observations; $X_{ji} \in \mathbb{R}^{d_j}$ are regressors in the $j$th equation, $m_j$; $j = 1, \ldots, P$, are unknown smooth regression functions; and $U_i = (U_{1i}, \ldots, U_{Pi})'$ are random disturbances such that $E\left(U_{ji} \mid X_{ji}\right) = 0$ and $E(U_i U_i' \mid X_i) = \sum(X_i)$, with $X_i$ being the disjoint union of $X_{1i}, \ldots, X_{Pi}$.

We are interested in testing the null hypothesis that

$$H_0: \text{all off-diagonal elements of } \sum(X_i) \text{ are zero a.s.}; \tag{2.2}$$

i.e., the $U_{ij}$'s are conditionally uncorrelated given $X_i$. The alternative hypothesis is

$$H_1: \text{some of the off-diagonal elements of} \sum(X_i) \text{ are nonzero.} \tag{2.3}$$

For notational simplicity, we will focus on the case in which $p = 2$ and remark on the other cases in Section 3.4. In the former case, we write

$$\sum(x) = \begin{pmatrix} \sigma_1^2(x) & \sigma_1(x)\sigma_2(x)\rho(x) \\ \sigma_1(x)\sigma_2(x)\rho(x) & \sigma_2^2(x) \end{pmatrix}. \tag{2.4}$$

We are interested in testing whether $P(\rho(X) = 0) = 1$, where here and below the probability is taken over $X$. Alternatively, let $\sigma_{12}(x) = \sigma_1(x)\sigma_2(x)\rho(x)$. The null hypothesis is

$$H_0: P(\sigma_{12}(X) = 0) = 1; \tag{2.5}$$

and the alternative hypothesis is

$$H_1: P(\sigma_{12}(X) = 0) < 1. \tag{2.6}$$

Let $f(x)$ denote the density function of $X_i$. When the null hypothesis in (5) is written as $E(U_1U_2|X) = 0$ a.s., we can construct consistent tests of $H_0$ versus $H_1$ using various distance measures. A convenient choice is to use the measure

$$\Gamma = \int \sigma_{12}^2(x) f^2(x) dx \geq 0 \tag{2.7}$$

and $\Gamma = 0$ if and only if $H_0$ given by (5) holds. Note that the use of density weight in the definition of $\Gamma$ will help us avoid the random denominator issue. We will propose a test statistic based upon a kernel estimator of $\Gamma$.

### 2.2 The Test Statistic

The proposed test is based upon an estimator of the conditional covariance between $U_{1i}$ and $U_{2i}$ given $X_i$. For the moment, assume that we observe the sequence $\{U_{1i}, U_{2i}\}$ together with $\{X_i\}$, where $X_i \in \mathbb{R}^d$ is a disjoint union of $X_{1i}$ and $X_{2i}$. Under the additional assumption that $E(U_{ji}|X_i) = 0$, $j = 1, 2$, we can estimate the conditional covariance of $U_{1i}$ and $U_{2i}$ given $X_i$ by

$$\hat{\sigma}_{12}(x) = \frac{\frac{1}{n}\sum_{i=1}^n K_H(x - X_i)U_{1i}U_{2i}}{\frac{1}{n}\sum_{i=1}^n K_H(x - X_i)}, \tag{2.8}$$

where $K$ is a symmetric kernel function defined on $\mathbb{R}^d$, $H = \mathrm{diag}(h_1, \ldots, h_d)$ is a matrix of bandwidth sequences, and $K_H(u) = |H|^{-1}K(H^{-1}u)$ with $|H|$ being the determinant of $H$. Further, denote

$$\hat{f}(x) = \frac{1}{n}\sum_{i=1}^n K_H\left(x - X_i\right) \tag{2.9}$$

the nonparametric kernel estimate of the density function $f(x)$ of $X_i$.

We then estimate $\Gamma$ by the following functional:

$$\begin{aligned}\Gamma_{1n} &= \int \left[\hat{\sigma}_{12}(x)\right]^2 \hat{f}^2(x)\, dx \\ &= \frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n \tilde{U}_{1i}\tilde{U}_{2i}\tilde{U}_{1j}\tilde{U}_{2j}\overline{K}_H\left(X_j - X_i\right),\end{aligned} \tag{2.10}$$

where $\overline{K}_H(u) = |H|^{-1}\overline{K}(H^{-1}u)$ and $\overline{K}(u) = \int K(v)K(u - v)\, dv$ is the convolution kernel derived from $K$. For example, if $K(u) = \exp(-u^2/2)/\sqrt{2\pi}$, then $\overline{K}(u) = \exp(-u^2/4)/\sqrt{4\pi}$, a normal density with zero mean and variance 2.

The preceding statistic is simple to compute and offers a natural way to test $H_0$ in (5). Nevertheless, we propose a bias-adjusted test statistic; namely,

$$\Gamma_n = \frac{1}{n^2}\sum_{i=1}^n\sum_{j\neq i}^n U_{1i}U_{2i}U_{1j}U_{2j}\overline{K}_H\left(X_j - X_i\right). \tag{2.11}$$

In effect, $\Gamma_n$ removes the "diagonal" ($j = i$) terms from $\Gamma_{1n}$ in (10), thus reducing the bias of the statistic. A similar idea has been used in Lavergne and Vuong (2000) and Su and White (2007).

*When $\{U_{1i}, U_{2i}\}$ need to be estimated.* When $\{U_{1i}, U_{2i}\}$ are unobserved, which is the typical case in practice, we need to estimate them by the nonparametric residuals obtained from the regression of $Y_{ji}$ on $X_{ji}$, $j = 1, 2$. In this article, we give asymptotic analysis based on the local polynomial procedure. See Fan and Gijbels (1996) for discussions on the attractive properties of the local polynomial estimator.

Let $K_1$ and $K_2$ denote kernel functions on $\mathbb{R}^{d_1}$ and $\mathbb{R}^{d_2}$, respectively. Let $H_j = \text{diag}\,(h_{j1}, \ldots, h_{jd_j})$ and $K_{H_j}(z) = |H_j|^{-1} K_j\left(H_j^{-1}z\right), j = 1, 2$. For the dataset $\{Y_{ji}, X_{ji}\}_{i=1}^n$, the $p_j$th order local polynomial estimate of $m_j(x_j)$ can be obtained from the weighted least squares regression:

$$\tilde{\theta}^{(j)}(x_j) \equiv \underset{\theta^{(j)}}{\arg\min} \sum_{t=1}^{n} K_{H_j}((x_j - X_{ji})) \left[ Y_{ji} - \sum_{0 \le |\mathbf{i}| \le p_j} \theta_{\mathbf{i}}^{(j)} (X_{ji} - x_j)^{\mathbf{i}} \right]^2 . \tag{2.12}$$

Here, we use the notation of Masry (1996): $\mathbf{i} = (i_1, \ldots, i_{d_j})$, $|\mathbf{i}| = \sum_{l=1}^{d_j} i_l$, $z^{\mathbf{i}} = \Pi_{l=1}^{d_j} z_l^{i_l}$, $\sum_{0 \le |\mathbf{i}| \le p_j} = \sum_{l=0}^{p_j} \sum_{\substack{j_1=0 \cdots \\ j_1+\ldots+j_{d_1}=l}}^{l} \sum_{j_d=0}^{l}$, and $\theta^{(j)} = \theta^{(j)}(x_j)$ is a collection of all the parameters $\theta_{\mathbf{i}}^{(j)} = \theta_{\mathbf{i}}^{(j)}(x_j)$, $0 \le |\mathbf{i}| \le p_j$, in lexicographical order. We denote the first element of $\widetilde{\theta}^{(j)}(x_j)$ as $\tilde{m}_j(x_j)$, the $p_j$th order local polynomial estimate of $m_j(x_j)$.

We then estimate $U_{ji}$ by

$$\tilde{U}_{ji} = Y_{ji} - \tilde{m}_j(X_{ji}). \tag{2.13}$$

We replace $U_{ji}$ in $\Gamma_n$ by $\tilde{U}_{ji}$ to obtain the following test statistic:

$$\tilde{\Gamma}_n = \frac{1}{n^2} \sum_{i=1}^{n} \sum_{l \ne i}^{n} \tilde{U}_{1i} \tilde{U}_{2i} \tilde{U}_{1l} \tilde{U}_{2l} \overline{K}_H (X_l - X_i). \tag{2.14}$$

We will show that, after being appropriately scaled, $\tilde{\Gamma}_n$ is asymptotically normally distributed under suitable assumptions.

## 3. THE ASYMPTOTIC DISTRIBUTIONS OF THE TEST STATISTIC

In this section, we first present a set of assumptions that are used in deriving the asymptotic distributions of our test statistic. Then we study the asymptotic distribution of our test under both the null hypothesis and a sequence of local alternatives, followed by some remarks.

### 3.1 Assumptions

We make the following assumptions on the error terms, regressors, kernel functions, and bandwidth sequences.

*Assumptions*

*A1.* $\{X_i, U_{1i}, U_{2i}\}$ is a strictly stationary strong mixing process with mixing coefficients $\alpha(j)$ such that $\sum_{j=1}^{\infty} j^2 \alpha(j)^{\delta/(\delta+1)} < \infty$ for some constant $\delta \in (0, 1)$.

*A2.* For $j = 1, 2$, $E\left(U_{ji}|X_i, \ldots, X_1, U_{i-1}, \ldots, U_1\right) = 0$, $E\left(|\varepsilon_i|^{4(1+\delta)}\right) \le C < \infty$, and $E\left|\varepsilon_{i_1}^{t_1} \varepsilon_{i_2}^{t_2} \cdots \varepsilon_{i_l}^{t_l}\right|^{1+\delta} \le C < \infty$, where $\varepsilon_i \equiv U_{1i} U_{2i}$, $2 \le l \le 4$, $0 \le t_s \le 4$, and $\sum_{s=1}^{l} t_s \le 8$.

*A3.* (1) Let $\sigma^2(x) \equiv E\left(\varepsilon_i^2 \,|\, X_i = x\right)$ and $\mu_4(x) = E(\varepsilon_i^4 \,|\, X_i = x)$. Then both $\sigma^2(x)$ and $\mu_4(x)$ satisfy the Lipschitz condition: $|g(x + x^*) - g(x)| \le D(x)\|x^*\|$, where $g(\cdot) = \sigma^2(\cdot)$ or $\mu_4(\cdot)$, and $\|\cdot\|$ denotes the Euclidean norm. (2) For each $1 < i_1 < \ldots < i_l$ ($l = 1, 2, 3$), the joint density $f_{i_1,\ldots,i_l}(\cdot)$ of $(X_1, X_{i_1}, \ldots, X_{i_l})$ exists and satisfies the Lipschitz condition: $\left|f_{i_1,\ldots,i_l}\left(x^{(1)} + v^{(1)}, \ldots, x^{(l+1)} + v^{(l+1)}\right) - f_{i_1,\ldots,i_l}\left(x^{(1)}, \ldots, x^{(l+1)}\right)\right| \le D_{i_1,\ldots,i_l}\left(x^{(1)}, \ldots, x^{(l+1)}\right)\|v\|$, where $v' = (v^{(1)\prime}, \ldots, v^{(l+1)\prime})$, $D_{i_1,\ldots,i_l}(x^{(1)}, \ldots, x^{(l+1)})$ is integrable and satisfies the conditions: $\int D_{i_1,\ldots,i_l}\left(v^{(1)}, \ldots, v^{(l+1)}\right)\|v\|^{2\theta} dv < C < \infty$, $\int D_{i_1,\ldots,i_l}\left(v^{(1)}, \ldots, v^{(l+1)}\right)$, and $f_{i_1,\ldots,i_l}\left(v^{(1)}, \ldots, v^{(l+1)}\right) dv < C < \infty$ for some $\theta > 1$ and constant $C > 0$. (3) The marginal density $f(\cdot)$ of $X_i$ is Lipschitz continuous: $|f(x) - f(x^*)| \le C\, \|x - x^*\|$ for some $C > 0$.

*A4.* $m_j(x_j)$ has $(p_j + 1)$th partial derivatives that are bounded and Lipschitz continuous of order 1 on the compact support $\mathrm{X}_j$ of $X_{ji}$, where $p_j \ge 1$ for $j = 1, 2$.

*A5.* The kernel functions $K$, $K_1$, and $K_2$ are a product of univariate kernels $k$. The $k$ is a bounded, symmetric, and uniformly continuous density function such that $\int |u|^{2p(2+\delta)} k(u)\, du < \infty$, where $p = \max(p_1, p_2)$.

*A6.* Let $\eta_j = \|H_j\|^{p_j+1} + n^{-1/2}|H_j|^{-1/2}\sqrt{\log n}$ for $j = 1, 2$. As $n \to \infty$ (1) $n|H_j|^2/(\log n)^3 \to \infty$, $n\|H_j\|^{2(p_j+2)} \to 0$; (2) $\|H\| \to 0$, $n|H_j|\,|H|(\log n)^2 \to \infty$; and (3) $n|H|^{1/2}\ \eta_1^2 \eta_2^2 \to 0$, where, for example, $\|H_j\| = \{\text{tr}\,(H'_j H_j)\}^{1/2}$.

Assumptions A1–A3 are common in nonparametric estimation with strong mixing data (Masry 1996, Gao and King 2003). They are mainly used in the proof of Lemma A1 in the appendix. The smoothness condition in Assumption A4 and the assumptions on the kernels in Assumption A5 are typical in the literature with regard to local polynomial estimation. By Masry (1996), $\tilde{m}_j(x_j) - m_j(x_j) = O_p(\eta_j)$ uniformly in $x_j$. Even though we conjecture that Assumption A6 can be relaxed with a much more complicated proof of our main result in the next section, it is easily satisfied in many applications. For example, if we set $h_{ji} \propto n^{-1/\alpha_j}$ for $i = 1, \ldots, d_j$ and $j = 1, 2$, and set $h_i \propto n^{-1/\alpha}$ for $i = 1, \ldots, d$, then Assumption A6(1) requires that $\alpha_j \in \left(2d_j, 2(p_j + 2)\right)$ and Assumption A6(2) requires that $\alpha > \alpha_j d/(\alpha_i - d_j)\ \forall j$. Let

$$c_1 \equiv \alpha_1 \alpha_2 - 4(p_1 + 1)(p_2 + 1)$$

and

$$c_2 \equiv d_1/\alpha_1 + d_2/\alpha_2 - 1. \tag{3.1}$$

Given the choice of $p_j$ and $\alpha_j$, Assumption A6(3) implies that we can choose $\alpha$ such that

$$\begin{cases} \frac{\alpha}{d} > \max_j \left\{\frac{\alpha_j}{\alpha_j - d_j}\right\} & \text{if } c_1 \le 0 \text{ and } c_2 \le 0 \\ \max_j \left\{\frac{\alpha_j}{\alpha_j - d_j}\right\} < \frac{\alpha}{d} < \frac{\alpha_1\alpha_2}{2c_1} & \text{if } c_1 > 0 \text{ and } c_2 \le 0 \\ \max_j \left\{\frac{\alpha_j}{\alpha_j - d_j}\right\} < \frac{\alpha}{d} < \frac{1}{2c_2} & \text{if } c_2 > 0 \text{ and } c_1 \le 0 \\ \max_j \left\{\frac{\alpha_j}{\alpha_j - d_j}\right\} < \frac{\alpha}{d} < \min\left(\frac{\alpha_1\alpha_2}{2c_1}, \frac{1}{2c_2}\right) & \text{if } c_1 > 0 \text{ and } c_2 > 0 \end{cases} . \tag{3.2}$$

So if $c_1 \le 0$ and $c_2 \le 0$ in (3.2), we can choose $\alpha$ sufficiently large such that the convergence rate of $\|H\|$ to zero can be arbitrarily slow.

## 3.2 Asymptotic Null Distribution

To state our main result, recall $\sigma^2(x) = E(U_{1i}^2 U_{2i}^2 \mid X_i = x)$. Define $\sigma_0^2 \equiv 2\int \overline{K}^2(u)du\ E[\sigma^4(X_i)f(X_i)]$. Our main result is stated in Theorem 3.1.

*Theorem 3.1.* Under Assumptions A.1–A.6 and under $H_0$, $n|H|^{1/2}\tilde{\Gamma}_n \xrightarrow{d} N(0, \sigma_0^2)$.

The proof is tedious and is relegated to the appendix. From the proof, we know that $n|H|^{1/2}\tilde{\Gamma}_n = n|H|^{1/2}\Gamma_n + o_p(1)$, which means that the first-stage estimation of the conditional mean functions does not affect the first-order asymptotic properties of the test. To implement the test, we require a consistent estimate of the variance $\sigma_0^2$. Define

$$\hat{\sigma}^2 \equiv 2n^{-2}|H| \sum_{i=1}^{n} \sum_{j\neq i}^{n} \tilde{U}_{1i}^2 \tilde{U}_{2i}^2 \tilde{U}_{1j}^2 \tilde{U}_{2j}^2 \overline{K}_H^2 (X_i - X_j). \tag{3.3}$$

It is easy to show that $\hat{\sigma}^2$ is consistent for $\sigma_0^2$ under $H_0$. We then compare

$$T_n \equiv n|H|^{1/2}\tilde{\Gamma}_n / \sqrt{\hat{\sigma}^2}, \tag{3.4}$$

with the one-sided critical value $z_\alpha$ from the standard normal distribution, and reject the null when $T_n > z_\alpha$.

## 3.3 Asymptotic Local Power Property

To examine the asymptotic local power of our test, we consider the following local alternatives:

$$H_1(\gamma_n) : \sigma_{12}(x) = \gamma_n \Delta(x), \tag{3.5}$$

where $\Delta(x)$ satisfies $E|\Delta(X)|^{2+\delta} < \infty$. Define

$$\Delta_0 \equiv \int \Delta^2(x) f^2(x) dx. \tag{3.6}$$

The following proposition shows that our test can distinguish local alternatives $H_1(\gamma_n)$ at rate $\gamma_n = n^{-1/2}|H|^{-1/4}$ while maintaining a constant level of asymptotic power.

*Proposition 3.2.* Under Assumptions A.1–A.6, suppose that $\gamma_n = n^{-1/2}|H|^{-1/4}$ in $H_1(\gamma_n)$. Then, the power of the test satisfies $P(T_n \geq z_\alpha | H_1(\gamma_n)) \to 1 - \Phi(z_\alpha - \Delta_0/\sigma_0)$, where $\Phi(\cdot)$ is the cumulative distribution function of standard normal.

*Remark.* Proposition 3.2 implies that the test has nontrivial asymptotic power against alternatives for which $\Delta_0 \equiv \int \Delta^2(x) f^2(x) dx > 0$. The power increases with the magnitude of $\Delta_0 / \sigma_0$. Furthermore, by taking a large bandwidth, we can make the alternative magnitude against which the test has nontrivial power, i.e., $\gamma_n$, arbitrarily close to the parametric rate $n^{-1/2}$. As remarked earlier on, this is possible if we can choose $p_j$ and $\alpha_j$ such that $c_j \leq 0$ in (3.1).

## 3.4 Remarks

Theorem 3.1 covers the asymptotic null distribution of the test statistic that is based on residuals resulting from nonparametric regression with continuous variables. While this case suffices for many empirical applications when the functional forms of the regression functions in a SUR or VAR system are unknown and only continuous variables appear in the regression, our testing procedure is potentially applicable to a much wider range of situations. We now discuss several of these.

*3.4.1 Parametric SUR or VAR Models.* When the conditional mean functions in a SUR or VAR system are parametrically specified, we can first estimate the parametric model for each equation to obtain the residuals and then construct the test for conditional uncorrelatedness based upon these residuals. In this case, we do not need to specify conditions on the kernels $K_j$ and bandwidths $H_j$ $(j = 1, 2)$ in Assumptions A5 and A6. The asymptotic results in Theorem 3.1 and Proposition 3.2 follow with a much simpler proof than the current case. To see why, notice that the unknown finite dimensional parameters can be estimated at the regular $n^{-1/2}$-rate, which implies that $\tilde{U}_{ji} = U_{ji} + O_p(n^{-1/2})$, where $\tilde{U}_{ji}$ is now the residual obtained from the parametric regression. As a result, the first-stage parametric estimation does not have any asymptotic impact on the first-order asymptotic property of the test given the fact that $\|H\| \to 0$ as $n \to \infty$.

*3.4.2 Test for More Than One Pair of Conditional Uncorrelatedness* Our test can easily be extended to test for more than one pair of conditional uncorrelatedness. Let $\sigma_{kl}(x)$ be the $(k, l)$th element of $\sum(x)$ for $k, l = 1, \ldots, P$, where $\sum(x) = E(U_i U'_i | X_i = x)$ was defined after (2.1). Now the functional $\Gamma$ in (2.7) becomes

$$\Gamma = \sum_{k=1}^{P-1} \sum_{l=k+1}^{P} \int \sigma_{kl}^2(x) f^2(x) dx \tag{3.7}$$

and the test statistic $\tilde{\Gamma}_n$ in (2.14) becomes

$$\tilde{\Gamma}_n = \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j\neq i}^{n} \sum_{k=1}^{P-1} \sum_{l=k+1}^{P} \tilde{U}_{ki} \tilde{U}_{li} \tilde{U}_{kj} \tilde{U}_{lj} \overline{K}_H (X_j - X_i), \tag{3.8}$$

where $\tilde{U}_{ki}$ is as defined in (2.13). Theorem 3.1 continues to hold by replacing $\sigma^2(x)$ in the definition of ${\sigma_0}^2$ by

$$\sigma^2(x) \equiv E\left[ \left( \sum_{k=1}^{P-1} \sum_{l=k+1}^{P} U_{ki} U_{li} \right)^2 \middle| X_i = x \right]. \tag{3.9}$$

*3.4.3. Connection with Tests for Unconditional Correlation in Parametric Models* Breusch and Pagan (1980) proposed a Lagrange multiplier (LM) statistic for testing zero cross-equation correlations in the parametric SUR system:

$$Y_{ji} = \beta'_j X_{ji} + U_{ji}, \quad j = 1, \ldots, P, \quad i = 1, \ldots, n, \tag{3.10}$$

where $\beta_j$'s are unknown regression coefficients. Their test is based on the following LM statistic:

$$\mathrm{LM}_n = n \sum_{k=1}^{P-1} \sum_{l=k+1}^{P} \hat{\rho}_{kl}^2, \tag{3.11}$$

where $\hat{\rho}_{kl}$ is the sample analog estimate of the pairwise correlation of the residuals $\tilde{U}_{ki}$ and $\tilde{U}_{li}$, which are obtained from the OLS regression of $Y_{ji}$ on $X_{ji}$, $j = k, l$. Under the null hypothesis,

$$H_0^* : \mathrm{cov}(U_{k1}, U_{l1}) = 0 \quad \text{for all } k \neq l, \tag{3.12}$$

$\mathrm{LM}_n \xrightarrow{d} \chi^2(P(P-1)/2)$ as $n \to \infty$, and $P$ is held fixed. As it stands, this test is not applicable when $P \to \infty$.

Noting that, under $H_0^*$, $H_0^*$, $n\hat{\rho}_{kl}^2 \xrightarrow{d} \chi^2(1)$ as $n \to \infty$ and $\hat{\rho}_{kl}$'s, $k \neq l$, are asymptotically uncorrelated, Pesaran, Ullah, and Yamagata (2006) considered a bias-adjusted normal approximation version of the LM statistic:

$$\mathrm{NLM}_n = \sqrt{\frac{1}{P(P-1)}}\sum_{k=1}^{P-1}\sum_{l=k+1}^{P}(n\hat{\rho}_{kl}^2 - 1), \tag{3.13}$$

which is asymptotically distributed $N(0,1)$ under $H_0^*$ when $n \to \infty$ first and then $P \to \infty$. Since $\mathrm{NLM_n}$ is likely to exhibit substantial size distortions for $P$ large and $n$ small, these authors also proposed two mean-biased-adjusted versions of the $\mathrm{NLM_n}$ test statistic. These tests are consistent even in the situations where the Pesaran's (2004) Cross-section Dependence test is inconsistent.

To see the connection of our test with the $\mathrm{NLM_n}$ test, suppose that each element of the bandwidth matrix $H$ tends to $\infty$ as $n \to \infty$; then for $\tilde{\Gamma}_n$ in (3.8) we have

$$\tilde{\Gamma}_n \simeq \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\sum_{k=1}^{P-1}\sum_{l=k+1}^{P}\tilde{U}_{ki}\tilde{U}_{li}\tilde{U}_{kj}\tilde{U}_{lj}\overline{K}_H(0) \tag{3.14}$$

where $a_n \simeq b_n$ signifies that $a_n = b_n\big(1 + o_p(1)\big)$. Hence, in this case, $\tilde{\Gamma}_n$ will reduce to calculating the sum of squared unconditional covariances with bias correction by excluding $j = i$ terms in the summation in (3.14). Consequently, our test can be regarded as the conditional analog of the $\mathrm{NLM_n}$ test. The difference is that the $\mathrm{NLM_n}$ test is based on the estimate of unconditional correlation, whereas our test is based on the estimate of conditional covariance. Also, the $\mathrm{NLM_n}$ is developed for the parametric regression model, whereas the regression model in our test is nonparametric.

*3.4.4. Testing for conditional homoscedasticity* In some sense, the test for conditional homoscedasticity can be regarded as a special case of our test. To see this, consider $\sigma_1^2(x_1) = E\big(U_{1i}^2|X_{1i} = x_1\big)$. Under the null of conditional homoscedasticity, $P\big(\sigma_1^2(X_{1i}) = \bar{\sigma}_1^2\big) = 1$ for some $\bar{\sigma}_1^2 > 0$. Hall and Carroll (1989) showed that $\bar{\sigma}_1^2$ can be estimated at the parametric $n^{-1/2}$ rate even though the conditional mean function is only estimated at a nonparametric rate. Let $\tilde{\sigma}_1^2 = (1/n)\sum_{i=1}^{n}(Y_{1i} - \tilde{m}_1(X_{1i}))^2$. Define

$$\tilde{\Gamma}_{n1} = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\left(\tilde{U}_{1i}^2 - \tilde{\sigma}_1^2\right)\left(\tilde{U}_{1j}^2 - \tilde{\sigma}_1^2\right)\overline{K}_H\big(X_{1j} - X_{1i}\big). \tag{3.15}$$

A consistent test for conditional homoscedasticity can be based upon the following test statistic:

$$T_{n1} = n|H_1|^{1/2}\,\tilde{\Gamma}_{n1}/\sqrt{\hat{\sigma}_1^2} \tag{3.16}$$

where

$$\hat{\sigma}_1^2 \equiv 2n^{-2}|H_1|\sum_{i=1}^{n}\sum_{j\neq i}^{n}\left(\tilde{U}_{1i}^2 - \tilde{\sigma}_1^2\right)^2 \times\left(\tilde{U}_{1j}^2 - \tilde{\sigma}_1^2\right)^2\overline{K}_{H_1}^2\big(X_{1i} - X_{1j}\big). \tag{3.17}$$

Under some weak conditions and the null of conditional homoscedasticity, we can follow the proof of Theorem 3.1 to show that $T_{n1} \xrightarrow{d} N(0,1)$. The preceding argument goes through with little modification if we are interested in testing for constancy of conditional covariance:

$$H_0^*: P(\sigma_{12}(X_i) = \bar{\sigma}_{12}) = 1 \text{ for some } \bar{\sigma}_{12} \in \mathbb{R} \tag{3.18}$$

As a referee notes, our test for conditional homoscedasticity is very similar to that of Li and Hsiao (2001). The only difference is that we use the convolution kernel $\overline{K}_H(\cdot)$ in (3.15) instead of the original kernel $K(\cdot)$. The convolution kernel appears because in (2.10) we use $\hat{\sigma}_{12}(x)$ and $\hat{f}(x)$ in place of $\sigma_{12}(x)$ and $f(x)$ in (2.7). Let $\varepsilon_{1i} = U_{1i}^2 - \bar{\sigma}_1^2$ and $f_{X_1}(\cdot)$ be the density function of $X_{1i}$. Li and Hsiao's (2001) formula is different from ours, because they base their test upon the sample analog of $E[\varepsilon_{1i}E(\varepsilon_{1i}|X_{1i})f_{X_1}(X_{1i})]$, i.e., $n^{-1}\sum_{i=1}^{n}\varepsilon_{1i}E(\varepsilon_{1i}|X_{1i})f_{X_1}(X_{1i})$, which can be estimated by

$$\overline{\Gamma}_{n1} \equiv \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\left(\tilde{U}_{1i}^2 - \tilde{\sigma}_1^2\right)\left(\tilde{U}_{1j}^2 - \tilde{\sigma}_1^2\right)K_H\big(X_{1j} - X_{1i}\big). \tag{3.19}$$

## 4. MONTE CARLO SIMULATIONS

In this section, we conduct a small set of Monte Carlo simulations to evaluate the finite sample performance of our test.

### 4.1 Data Generating Processes

We first generate data from the following data generating processes (DGPs):

$$\mathrm{DGP1}: \begin{cases} Y_{1i} = \sin(8\pi X_{1i}) + U_{1i}, \\ Y_{2i} = \phi(X_{2i}, 0.2, 0.5) + \phi(X_{2i}, 0.8, 0.25) + U_{2i}, \end{cases}$$

where $\phi(x,a,b)$ is the normal density function with mean $a$ and standard deviation $b$; $U_{1i} = \sigma(X_{1i},\, X_{2i})V_{1i}$, $U_{2i} = \sigma(X_{1i}, X_{2i})V_{2i}$, $\sigma(x_1, x_2) = \sqrt{\exp(x_1 + x_2)}$, $\{X_{1i}\}$, and $\{X_{2i}\}$ are mutually independent iid $U(0,1)$:

$$\begin{pmatrix} V_{1i} \\ V_{2i} \end{pmatrix} \sim \text{iid } N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right), \tag{4.1}$$

and the process $\{X_i = (X_{1i}, X_{2i})\}$ is independent of the process $\{V_{1i}, V_{2i}\}$.

$$\text{DGP 2:} \begin{cases} Y_{1i} = \cos(2\pi X_{1i}) + U_{1i}, \\ Y_{2i} = \cos(2\pi X_{2i}) + U_{2i}, \end{cases}$$

where $U_{1i} = \sqrt{h_{1,i}}V_{1i}$, $U_{2i} = \sqrt{h_{2,i}}V_{2i}$, and $h_{j,i} = 0.5 + 0.1U_{j,i-1}^2 + 0.6h_{j,i-1} + 0.2X_i^2$ for $j = 1, 2$, $X_i = X_{1i} = X_{2i}$, and $X_i \sim$ iid $U$ (0.1); the distribution of $(V_{1i}, V_{2i})$ is given in (4.1), and the process $\{X_i\}$ is independent of the process $\{V_{1i}, V_{2i}\}$.

DGP 1 specifies an iid sequence $\{Y_{1i}, Y_{2i}, X_{1i}, X_{2i}\}$. DGP 2 yields a bivariate GARCH-$X$ process. In both cases, we will consider five different choices of $\rho$: 0, 0.3, 0.6, −0.3, and −0.6. When $\rho = 0$, we study the size behavior of the test. To conduct our test for DGP 2, we throw away the first 100 observations when generating the data.

In the next DGP, we consider linear specifications for the conditional mean and time-varying conditional correlation function:

$$\text{DGP 3: } \begin{cases} Y_{1i} = 1 + X_{1i} + U_{1i}, \\ Y_{2i} = 1 + 0.5X_{2i} + U_{2i}, \end{cases}$$

where $\{X_i = X_{1i} = X_{2i}\}$ are the iid sum of 48 independent random variables, each uniformly distributed on $[-0.25,0.25]$. According to the central limit theorem, we can treat $X_i$'s as being nearly standard normal random variables but with compact support $[-12, 12]$. The disturbance terms are generated as

$$\begin{pmatrix} U_{1i} \\ U_{2i} \end{pmatrix} | X_i \sim \text{iid } N\left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \sigma^2(X_i) \begin{pmatrix} 1 & \rho_i \\ \rho_i & 1 \end{pmatrix} \right), \tag{4.2}$$

where $\sigma^2(X_i) = 0.25 + X_i^2$, and $\rho_i = \rho \exp(X_i)/(1 + \exp(X_i))$. To save time, we will only consider three choices of $\rho$: 0, 0.6, and $-0.6$. When $\rho = 0$, we study the size behavior of the test. When $\rho = \pm 0.6$, given the fact that $X_i$ is approximately distributed as $N(0, 1)$, we can verify that $E(\rho_i) \approx \pm 0.3$.

Like Dufour and Khalaf (2002), who considered finite-sample procedures for testing no *correlation* in multiequation models, we will consider the test of no *conditional correlation* in a six-equation model in DGP 4.

DGP 4: $Y_{ji} = \alpha_{j1}X_{1i} + \alpha_{j2}X_{2i} + U_{ji}$, where $X_{1i}$, $X_{2i}$ are independently generated as $X_i$ in DGP 3,

$$\alpha_{j1} = 1 \text{ for } j = 1,2,3,4, \text{and } \alpha_{j1} = 0.5 \quad \text{for } j = 5,6,$$

$$\alpha_{j2} = 1 \text{ for } j = 1,2, \text{and } \alpha_{j2} = 0.5 \quad \text{for } j = 3,4,5,6,$$

$$U_{ji} = \sqrt{h_{j,i}} V_{ji}, V_i = (V_{1i}, \ldots, V_{6i})', X_i = (X_{1i}, X_{2i})',$$

$$V_i | X_i \sim \text{iid } N(0, A_i' A_i).$$

$$A_i = \sigma(X_i) \begin{pmatrix} 1 & \rho_{1i} & \rho_{2i} & \rho_{2i} & \rho_{2i} & \rho_{2i} \\ \rho_{1i} & 1 & \rho_{2i} & \rho_{2i} & \rho_{2i} & \rho_{2i} \\ \rho_{2i} & \rho_{2i} & 1 & \rho_{1i} & \rho_{2i} & \rho_{2i} \\ \rho_{2i} & \rho_{2i} & \rho_{1i} & 1 & \rho_{2i} & \rho_{2i} \\ \rho_{2i} & \rho_{2i} & \rho_{2i} & \rho_{2i} & 1 & \rho_{1i} \\ \rho_{2i} & \rho_{2i} & \rho_{2i} & \rho_{2i} & \rho_{1i} & 1 \end{pmatrix},$$

in which $h_{j,i} = 0.5 + 0.05U_{ji}^2 + \lambda_j h_{j,i-1}, \lambda_j = 0.9$ for $j = 1, 2$, $\lambda_j = 0.7$ for $j = 3, 4$, and $\lambda_j = 0.5$ for $j = 5, 6$; $\sigma(X_i) = 0.25 + X_{1i}^2 + X_{2i}^2, \rho_{1i} = 1.5\rho\cos(\pi i/10)$, and $\rho_{2i} = \rho\cos(\pi i/10)$. We will consider three choices of $\rho$: 0, 0.6 and $-0.6$. When $\rho = 0$, $A_i' A_i$ is a diagonal matrix so that we study the size behavior of the test. To conduct our test for DGP 4, we throw away the first 100 observations when generating the data.

## 4.2 Estimation and Bandwidth Choice

We will consider our test without and with estimation error. In the first case, we pretend that we observe $\{U_{ji}, j = 1, \ldots, P\}$ and test the conditional uncorrelatedness between $U_{ji}$'s given $X_i$. In the latter case, we first estimate the conditional mean function $m_j(\cdot)$ of $Y_{ji}$ given $X_{ji}$ at each data point to obtain the residual sequence $\{\tilde{U}_{ji}\}$, where $\tilde{U}_{ji} = Y_{ji} - \tilde{m}_j(X_{ji})$ with $\tilde{m}_j(x_j)$ being the local linear estimate of $m_j(x_j)$, $j = 1, \ldots, P$. Then we use $\{\tilde{U}_{ji}, j = 1, \ldots, P\}$ to test the conditional uncorrelatedness between $U_{ji}$'s given $X_i$.

Implementing the latter test requires methods for choosing two types of bandwidth parameters: $H_j$ ($j = 1, \ldots, P$) and $H$. The choice of the bandwidth sequences $\{H_j\}$ does not affect the rate at which our test can detect local alternatives as long as they satisfy Assumption A6, whereas the choice of $H$ does. So we now describe a systematic method for choosing $H$ and a rule of thumb for choosing $\{H_j\}$.

To estimate $m_j(\cdot)$ via the local linear regression, we might use the least squares cross-validated bandwidth $H_j^* = \text{diag}(h_{j1}^*, \ldots h_{jd_j}^*)$ that converges to zero at rate $n^{-1/(4+d_j)}$. For example, $d_j = 1$ for $j = 1, 2$ in DGPs 1-3; $d_j = 2$ for $j = 1, \ldots, 6$ in DGP 4. To eliminate the bias from the first-stage local linear regression, we follow Lee (2003, p.16) to use undersmoothing and adjust $h_{js}^*$ to

$$h_{js} = h_{js}^* n^{-1/20}, \quad s = 1, \ldots, d_j, \quad j = 1, \ldots, P.$$

We use the Gaussian kernel, $k(u) = \exp(-u^2/2)/\sqrt{2\pi}$, in all cases.

For the bandwidth matrix $H$, we set $H = \text{diag}(h_1, \ldots, h_d)$, where $d = 2$ in DGPs 1 and 4 and $d = 1$ in DGPs 2 and 3. Since it is difficult to pin down the optimal bandwidth for our test, we follow Horowitz and Spokoiny (2001, 2002) and consider a set of different bandwidth values of $H$. Like them, we use a geometric grid consisting of the points $h_{j,s} = \omega^s s_j h_{\min}$ ($s = 0, 1, \ldots, \mathrm{N} - 1; j = 1, \ldots, d$), where $s_j$ is the sample standard deviation of the $j$th element in $X_i$, N is the number of grid points, and $\omega = (h_{\max}/h_{\min})^{1/\mathrm{N}-1}$, with $h_{\min} = n^{-4d/3}$ and $h_{\max} = 4n^{-1/1{,}000}$. Following Horowitz and Spokoiny (2002), we choose N according to the rule of thumb $\mathrm{N} = [\log n] + 1$, where $[a]$ means the integer part of $a$. Let $H_s = \text{diag}(h_{1,s}, \ldots, h_{d,s})$, $s = 0, 1, \ldots, \mathrm{N} - 1$. For each $H_s$, we calculate the test statistic in (3.4) and denote it as $T_n(H_s)$. Define

$$\text{Sup}T_n \equiv \max_{0 \le s \le \mathcal{N}-1} T_n(H_s). \tag{4.3}$$

Even though $T_n(H_s)$ is asymptotically distributed as $N(0, 1)$ under the null for each $s$, the distribution of $\text{Sup}T_n$ is generally unknown. Fortunately, we can use bootstrap approximation.

The wild bootstrap proposed by Wu (1986) and Liu (1988) is designed to allow heteroscedasticity in the linear regression models. It has been examined in the time series context by Kreiss (1997), Hafner and Herwartz (2000), and Xu (2006), among others. We obtain the wild bootstrap residuals by

$$U_{ji}^* = \hat{U}_{ji} \upsilon_{ji}, \quad j = 1, \ldots, P, \quad i = 1, \ldots, n,$$

where $\hat{U}_{ji} = U_{ji}$ and $\tilde{U}_{ji}, \{\upsilon_{ji}\}$ are mutually independent iid sequences; they are independent of the process $\{X_i, Y_{1i}, \ldots, Y_{Pi}\}$, and $E(\upsilon_{ji}) = 0$, $E(\upsilon_{ij}^2) = 1$. There are many ways to obtain such sequences $\{\upsilon_{ji}, j = 1, \ldots, P\}$. In our simulation, we draw them independently from a distribution with probability masses $p = (1 + \sqrt{5})/(2\sqrt{5})$ and $1 - p$ at the points $(1 - \sqrt{5})/2$ and $(1 + \sqrt{5})/2$, respectively.

Based upon the bootstrap resampling data $\{U_{1i}^*, \ldots, U_{Pi}^*, X_i\}_{i=1}^n$, we construct the bootstrap version $\text{Sup}T_n^*$ of the test statistic $\text{Sup}T_n$. We repeat this procedure $B$ times and obtain the sequence $\left\{\text{Sup}T_{n,j}^*\right\}_{j=1}^B$. We reject the null when $p^* = B^{-1} \sum_{b=1}^B 1\left(\text{Sup}T_n \le \text{Sup}T_{n,b}^*\right)$ is smaller than the given level of significance, where $1(\cdot)$ is the usual indicator function.

Table 1. Finite sample rejection frequency: DGP 1

| Sample Size ($n$) | Correlation size ($\rho$) | No estimation 1% | No estimation 5% | No estimation 10% | With estimation 1% | With estimation 5% | With estimation 10% |
|---|---|---|---|---|---|---|---|
| 50 | 0 | 0.011 | 0.032 | 0.060 | 0.016 | 0.038 | 0.059 |
| | 0.3 | 0.234 | 0.395 | 0.471 | 0.178 | 0.306 | 0.396 |
| | 0.6 | 0.938 | 0.987 | 0.992 | 0.832 | 0.932 | 0.963 |
| | −0.3 | 0.257 | 0.422 | 0.496 | 0.131 | 0.295 | 0.375 |
| | −0.6 | 0.948 | 0.984 | 0.991 | 0.834 | 0.927 | 0.952 |
| 100 | 0 | 0.010 | 0.027 | 0.051 | 0.013 | 0.033 | 0.062 |
| | 0.3 | 0.536 | 0.678 | 0.736 | 0.453 | 0.600 | 0.677 |
| | 0.6 | 1 | 1 | 1 | 1 | 1 | 1 |
| | −0.3 | 0.555 | 0.696 | 0.752 | 0.475 | 0.606 | 0.671 |
| | −0.6 | 1 | 1 | 1 | 1 | 1 | 1 |
| 200 | 0 | 0.006 | 0.031 | 0.050 | 0.008 | 0.035 | 0.065 |
| | 0.3 | 0.912 | 0.960 | 0.977 | 0.866 | 0.937 | 0.954 |
| | 0.6 | 1 | 1 | 1 | 1 | 1 | 1 |
| | −0.3 | 0.885 | 0.945 | 0.966 | 0.848 | 0.922 | 0.947 |
| | −0.6 | 1 | 1 | 1 | 1 | 1 | 1 |

### 4.3 Simulation Results

We use 1,000 replications for each DGP and 399 bootstrap resamples in each replication. Tables 1–4 report the finite sample performance of our test. Those rows corresponding to $\rho = 0$ indicate how the size of our test behaves in each DGP.

We first summarize some important findings from Tables 1 and 2. First, we observe that our test is undersized in DGP 1 for each sample size under consideration and behaves reasonably well for DGP 2. Despite this, we find that the power of our test is great. When $\rho = \pm 0.3$, the conditional correlatedness between the two disturbance processes is weak. Yet Tables 1 and 2 suggest that our test is fairly powerful for such small sample size as $n = 100$. As $n$ grows, the power of the test also increases quickly. When $\rho = \pm 0.6$, our test even works very well for sample sizes as small as 50. Second, in all cases, we observe that the first-stage local linear estimation of the conditional mean functions does not have much effect on the size and power performance of the test. Third, the behavior of the test seems to be symmetric in $\rho$, which is in support of the local power property of our test.

Even though they are not reported here, we also did simulations on the bootstrap test based upon $T_n$ with a single choice of $H$: $H = \mathrm{diag}(h_1, \ldots, h_d)$, where $h_1 = cs_l n^{-1/(4+d)}$, with $l = 1, \ldots, d$, and $c = 0.5, 1, 2, 4$. We find that the choice of $c$, and thus the bandwidth $H$, does not have a significant impact on the level of the test, but it does influence the power of the test. The larger the value of $c$, the larger is the power. This supports our observation in the remark after (3.2a) and suggests that, in practice, we can choose a relatively large value of $c$.

Table 3 reports the simulation results for DGP 3. Notice that the conditional correlation $\rho_i$ between the two disturbance sequences varies according to $X_i$, and the parameter $\rho$ in Table 3 does not indicate the conditional correlation between the two

Table 2. Finite sample rejection frequency: DGP 2

| Sample Size ($n$) | Correlation size ($\rho$) | No estimation 1% | No estimation 5% | No estimation 10% | With estimation 1% | With estimation 5% | With estimation 10% |
|---|---|---|---|---|---|---|---|
| 50 | 0 | 0.011 | 0.041 | 0.074 | 0.018 | 0.059 | 0.094 |
| | 0.3 | 0.302 | 0.476 | 0.574 | 0.305 | 0.480 | 0.570 |
| | 0.6 | 0.975 | 0.999 | 1 | 0.959 | 0.989 | 0.995 |
| | −0.3 | 0.283 | 0.436 | 0.526 | 0.257 | 0.395 | 0.479 |
| | −0.6 | 0.958 | 0.990 | 0.996 | 0.929 | 0.978 | 0.986 |
| 100 | 0 | 0.011 | 0.040 | 0.072 | 0.017 | 0.048 | 0.078 |
| | 0.3 | 0.637 | 0.770 | 0.835 | 0.650 | 0.774 | 0.832 |
| | 0.6 | 1 | 1 | 1 | 1 | 1 | 1 |
| | −0.3 | 0.631 | 0.777 | 0.841 | 0.606 | 0.744 | 0.807 |
| | −0.6 | 1 | 1 | 1 | 1 | 1 | 1 |
| 200 | 0 | 0.008 | 0.039 | 0.071 | 0.007 | 0.040 | 0.076 |
| | 0.3 | 0.935 | 0.972 | 0.986 | 0.933 | 0.976 | 0.985 |
| | 0.6 | 1 | 1 | 1 | 1 | 1 | 1 |
| | −0.3 | 0.946 | 0.973 | 0.988 | 0.942 | 0.979 | 0.987 |
| | −0.6 | 1 | 1 | 1 | 1 | 1 | 1 |

Table 3. Finite sample rejection frequency: DGP 3

| Sample Size ($n$) | Correlation size ($\rho$) | No estimation 1% | 5% | 10% | With estimation 1% | 5% | 10% |
|---|---|---|---|---|---|---|---|
| 50 | 0 | 0.013 | 0.048 | 0.076 | 0.015 | 0.061 | 0.106 |
| | 0.6 | 0.296 | 0.467 | 0.573 | 0.297 | 0.477 | 0.570 |
| | −0.6 | 0.282 | 0.472 | 0.564 | 0.292 | 0.481 | 0.572 |
| 100 | 0 | 0.011 | 0.052 | 0.082 | 0.012 | 0.060 | 0.092 |
| | 0.6 | 0.628 | 0.785 | 0.841 | 0.634 | 0.794 | 0.849 |
| | −0.6 | 0.602 | 0.768 | 0.829 | 0.602 | 0.773 | 0.829 |
| 200 | 0 | 0.010 | 0.044 | 0.080 | 0.013 | 0.051 | 0.088 |
| | 0.6 | 0.938 | 0.970 | 0.980 | 0.927 | 0.965 | 0.981 |
| | −0.6 | 0.925 | 0.969 | 0.986 | 0.927 | 0.974 | 0.982 |

disturbance sequences. In fact, $E(\rho_i) \approx \pm 0.3$ when $\rho = \pm 0.6$. From Table 3, we observe that the test is well behaved in both size and power, and the power is comparable to the case where $\rho = \pm 0.3$ in Tables 1 and 2. Once again, we observe that the first-stage estimation of the conditional mean function has little effect on the test in terms of power, and that the behavior of the test is symmetric in $\rho$.

Table 4 reports the simulation results for DGP 4. Note that we have six equations in this case and the conditional correlation matrix between the six disturbance sequences varies according to time $i$. Since $d_j = d = 2$ for $j = 1, \ldots, 6$, and the sample size $n = 50$ is fairly small for the nonparametric estimation of two-dimensional conditional mean functions, we see that the first-stage estimation plays a role when $n = 50$. In particular, the size of the test with estimation error is a little bit inflated, and so is the power. We conjecture this is due to the small sample size in the local linear estimation of the two-dimensional conditional mean functions. In fact, as the sample size increases, we observe improvement of the test in terms of size. When $n = 200$, the tests with and without estimation error perform quite close to each other. Once again, we observe that the test is powerful and its behavior appears to be symmetric in $\rho$ when $n$ is not too small.

## 5. CONCLUDING REMARKS

In this article, we propose a nonparametric test for conditional uncorrelatedness in SUR and VAR systems. The test statistic is asymptotically normally distributed under the null and can detect local alternatives at the nonparametric rate $n^{-1/2}$ $|H|^{-1/4}$. Simulations indicate that the test works very well in finite samples.

In case we fail to reject the null, our test suggests no evidence of conditional correlatedness in the error terms so that we can estimate each equation in the system separately. On the other hand, in the case where the null is rejected, the question is how to explore the conditional correlation among the error terms to improve the asymptotic efficiency of the estimators by estimating the regression functions in the system jointly. See Ruckstuhl, Welsh, and Carroll (2000) and Welsh and Yee (2006) along this line of research. Another concern about our test is that we have implicitly assumed that the conditioning variable $X_i$ exhibits a density. This assumption may be untenable in practice. We conjecture that one can extend our test to allow $X_i$ to be a mixture of continuous and discrete variables by using the important tools developed by Racine and Li (2004). We leave these for future research.

## APPENDIX

### A.1 PROOF OF THE MAIN RESULTS

We use $C$ to signify a generic constant whose exact value may vary from case to case and $a'$ to denote the transpose of $a$. We write $a_n \simeq b_n$ to signify that $a_n = b_n\,(1 + o_p\,(1))$.

Table 4. Finite sample rejection frequency: DGP 4

| Sample Size ($n$) | Correlation size ($\rho$) | No estimation 1% | 5% | 10% | With estimation 1% | 5% | 10% |
|---|---|---|---|---|---|---|---|
| 50 | 0 | 0.007 | 0.041 | 0.091 | 0.011 | 0.074 | 0.136 |
| | 0.6 | 0.311 | 0.493 | 0.578 | 0.390 | 0.581 | 0.662 |
| | −0.6 | 0.354 | 0.507 | 0.593 | 0.437 | 0.589 | 0.670 |
| 100 | 0 | 0.012 | 0.059 | 0.116 | 0.012 | 0.066 | 0.117 |
| | 0.6 | 0.595 | 0.761 | 0.820 | 0.631 | 0.781 | 0.851 |
| | −0.6 | 0.594 | 0.752 | 0.816 | 0.641 | 0.784 | 0.843 |
| 200 | 0 | 0.009 | 0.057 | 0.104 | 0.012 | 0.064 | 0.117 |
| | 0.6 | 0.861 | 0.936 | 0.930 | 0.876 | 0.947 | 0.970 |
| | −0.6 | 0.862 | 0.930 | 0.957 | 0.866 | 0.943 | 0.969 |

### A.1.1 Proof of Theorem 3.1

Let $\delta_{jn}(x_j) = \tilde{m}_j(x_j) - m_j(x_j)$, $j = 1, 2$. Noting that $\tilde{U}_{ji} = U_{ji} - \delta_{jn}(X_{ji})$, we have

$$n^{-1}\sum_{i=1}^{n} \tilde{U}_{1i}\tilde{U}_{2i}K_H(x - X_i) = \sum_{j=1}^{4} A_{jn}(x), \quad \text{and}$$

$$\frac{1}{n^2}\sum_{i=1}^{n} \tilde{U}_{1i}^2\tilde{U}_{2i}^2\overline{K}_H(0) = \sum_{j=1}^{4} B_{jn},$$

where

$$A_{1n}(x) = \frac{1}{n}\sum_{i=1}^{n} U_{1i}U_{2i}K_H(x - X_i),$$

$$B_{1n} = \frac{1}{n^2}\sum_{i=1}^{n} U_{1i}^2U_{2i}^2\overline{K}_H(0),$$

$$A_{2n}(x) = \frac{1}{n}\sum_{i=1}^{n} \delta_{1n}(X_{1i})\delta_{2n}(X_{2i})K_H(x - X_i),$$

$$B_{2n} = \frac{1}{n^2}\sum_{i=1}^{n} \delta_{1n}^2(X_{1i})\delta_{2n}^2(X_{2i})\overline{K}_H(0),$$

$$A_{3n}(x) = -\frac{1}{n}\sum_{i=1}^{n} U_{1i}\delta_{2n}(X_{2i})K_H(x - X_i),$$

$$B_{3n} = \frac{1}{n^2}\sum_{i=1}^{n} U_{1i}^2\delta_{2n}^2(X_{2i})\overline{K}_H(0),$$

$$A_{4n}(x) = -\frac{1}{n}\sum_{i=1}^{n} U_{2i}\delta_{1n}(X_{1i})K_H(x - X_i),$$

$$B_{4n} = \frac{1}{n^2}\sum_{i=1}^{n} U_{2i}^2\delta_{1n}^2(X_{2i})\overline{K}_H(0).$$

Consequently,

$$\begin{aligned}
\tilde{\Gamma}_n &= \int \left[\frac{1}{n}\sum_{i=1}^{n}\tilde{U}_{1i}\tilde{U}_{2i}K_H(x - X_i)\right]^2 dx \\
&\quad - \frac{1}{n^2}\sum_{i=1}^{n}\tilde{U}_{1i}^2\tilde{U}_{2i}^2\overline{K}_H(0) \\
&= \int \Bigg\{\sum_{j=1}^{4} A_{jn}^2(x) + 2A_{1n}(x)A_{2n}(x) + 2A_{1n}(x)A_{3n}(x) \\
&\quad + 2A_{1n}(x)A_{4n}(x) + 2A_{2n}(x)A_{3n}(x) + 2A_{2n}(x)A_{4n}(x) \\
&\quad + 2A_{3n}(x)A_{4n}(x)\Bigg\}dx - \sum_{j=1}^{4} B_{jn}.
\end{aligned}$$

We complete the proof of the theorem by proving Lemmas A.1–A.10. ■

### Proof of Proposition 3.2

Under $H_1$ ($n^{-1/2}$ $|H|^{-1/4}$), we can verify that Lemmas A.2-A.10 continue to hold. Let $\epsilon_i = U_{1i}U_{2i}$. Let $E_{Xi}$ $(\epsilon_i)$ denote the conditional expectation of $\epsilon_i$ given $X_i$ and $\bar{\epsilon}_i = \epsilon_i - E_{Xi}(\epsilon_i)$. Then we can write $C_{1n} = C_{1n,a} + C_{1n,b} + C_{1n,c}$, where

$$C_{1n,a} = n^{-1}\sum_{i=1}^{n}\sum_{j\neq i}^{n} |H|^{1/2}\bar{\epsilon}_i\,\bar{\epsilon}_j\overline{K}_H(X_i - X_j),$$

$$C_{1n,b} = n^{-1}\sum_{i=1}^{n}\sum_{j\neq i}^{n} |H|^{1/2}E_{X_i}(\epsilon_i)E_{X_j}(\epsilon_j)\overline{K}_H(X_i - X_j), \text{ and}$$

$$C_{1n,c} = 2n^{-1}\sum_{i=1}^{n}\sum_{j\neq i}^{n} |H|^{1/2}\,\bar{\epsilon}_iE_{X_j}(\epsilon_j)\overline{K}_H(X_i - X_j).$$

Analogously to the proof of Lemma A.1, we can show that $C_{1n,a} \xrightarrow{d} N(0, \sigma_0^2)$. Next, noticing that $C_{1n,b} = n^{-2}\sum_{i=1}^{n}\sum_{j\neq i}^{n}$ $\Delta(X_i)\Delta(X_j)\overline{K}_H(X_i - X_j)$ is a second-order $U$ statistic, we can apply the LLN for the U-statistic of strong mixing process (e.g., Lee, 1990, p.122) to conclude that $C_{1n,b} \xrightarrow{p} \int \Delta^2(x)f^2(x)dx \equiv \Delta_0$. Now, write $C_{1n,c} = a_{1n}+b_{1n}$, where $a_{1n} = 2n^{-3/2}\sum_{1\le i<j\le n}^{n} |H|^{1/4}$ $\bar{\epsilon}_i\Delta(X_j)\overline{K}_H(X_i - X_j)$ and $b_{1n} = 2n^{-3/2}\sum_{1\le i<j\le n}^{n} |H|^{1/4}\bar{\epsilon}_j\Delta(X_i)$ $\overline{K}_H(X_j - X_i)$. By Assumption A2, $E(b_{1n})= 0$. By the Davydov's inequality (e.g., Bosq, 1998, p. 21), we have

$$\begin{aligned}
E(a_{1n}) &= 2n^{-3/2}|H|^{1/4} \\
&\quad \sum_{1\le i<j\le n}^{n} \int E\left[\bar{\epsilon}_iK_H(x - X_i)\Delta(X_j)K_H(x - X_j)\right]dx \\
&\le Cn^{-1/2}|H|^{1/4}|H|^{-\frac{2(1+\delta)}{2+\delta}}\sum_{i=1}^{n-1}[\alpha(i)]^{\delta/(2+\delta)} \\
&= o(1) \text{ for sufficiently small } \delta > 0.
\end{aligned}$$

Similarly, we can show that $E(a_{1n}^2) = o(1)$ and $E(b_{1n}^2) = o(1)$ Then $C_{1n,c} = o_p(1)$ by the Chebyshev inequality. Consequently, $P(T_n \ge z_\alpha|H_1(n^{-1/2}|H|^{-1/4})) \to 1 - \Phi(z_\alpha - \Delta_0/\sigma_0)$. ■

*Lemma A.1* Under the null, $C_{1n} \equiv n|H|^{1/2}\{\int A_{1n}^2(x)dx - B_{1n}\} \xrightarrow{d} N(0, \sigma_0^2)$

*Proof.* Let $\xi_i = (X'_i, U_{1i}, U_{2i})'$. Write $C_{1n} = 2n^{-1}\sum_{1\le i<j\le n}\phi(\xi_i, \xi_j)$, where $\phi(\xi_i, \xi_j) = |H|^{1/2}U_{1i}U_{2i}U_{1j}U_{2j}\overline{K}_H$ $(X_i - X_j)$. $C_{1n}$ is a second-order $U$ statistic and it is degenerate under the null. Under Assumptions A1, A2, A3(1) and (2), A5, and A6, one can verify the conditions of Lemma B.1 in Gao and King (2003) are satisfied so that a central limit theorem applies to $C_{1n}$. The asymptotic variance is given by $\operatorname{plim}_{n\to\infty}2E\phi$ $(\xi_0, \xi_1)^2 = \sigma_0^2$, where $\xi_0$ is an independent copy of $\xi_1$. ■

To proceed, let $f_j(x_j)$ denote the density of $X_{ji}$. By Masry (1996, Proposition 2, Theorems 4 and 6) and our symmetric assumption on the kernels, we have that, uniformly in $x_j \in \mathcal{X}_j$,

$$\begin{aligned}
\delta_{jn}(x_j) &= \frac{1}{nf_j(x_j)}\sum_{i=1}^{n} K_{H_j}(x_j - X_{ji})U_{ji} \\
&\quad + \frac{1}{nf_j(x_j)}\sum_{i=1}^{n} K_{H_j}(x_j - X_{ji})b_j(X_{ji}, x_j) \qquad \text{(A.1)} \\
&\quad + O_p\left(\left(\|H_j\| + n^{-1/2}|H_j|^{-1/2}\sqrt{\log n}\right)\eta_j\right) \\
&= O_p(\eta_j),
\end{aligned}$$

where $b_j(X_{ji}, x_j)$is the $(p_j + 1)$th polynomial function of $X_{ji} - x_j$ with coefficients given by $m_j^{(p_j+1)}(x_j)$, the $(p_j + 1)$th derivative of $m_j(x_j)$. For example, if $p_j = 1$, then $b_j(X_{ji}, x_j) = 1/2(X_{ji} - x_j)'\ddot{m}_j(x_j)(X_{ji} - x_j)$. Note that the result of Proposition 2 of Masry (1996) can be strengthened to $\sup_{x\in\mathbb{R}^d} E\{e_{n,\mathbf{j}}(x)\} = O(h_n^{p+2})$ using Masry's notation and our Lipschitz continuity condition on the $(p_j + 1)$th partial derivatives of $m_j(x)$, $j = 1, 2$.

*Lemma A. 2*

$$C_{2n} \equiv n|H|^{1/2}\left\{\int A_{2n}^2(x)dx - B_{2n}\right\} = o_p(1).$$

*Proof.* Noting that $\delta_{jn}(x_j) = O_p(\eta_j)$uniformly in $x_j \in \mathcal{X}_j$, we have

$$\begin{aligned}
C_{2n} &= n|H|^{1/2}\int\left[\frac{1}{n}\sum_{i=1}^{n}\delta_{1n}(X_{1i})\delta_{2n}(X_{2i})K_H(x - X_i)\right]^2 dx \\
&\quad - n|H|^{1/2}B_{2n} \\
&= n^{-1}|H|^{1/2}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\delta_{1n}(X_{1i})\delta_{2n}(X_{2i})\delta_{1n}(X_{1j})\delta_{2n}(X_{2j}) \\
&\quad \times \overline{K}_H(X_i - X_j) \\
&\leq Cn^{-1}|H|^{1/2}\max_{1\leq i\leq n}\delta_{1n}^2(X_{1i}) \\
&\qquad \max_{1\leq i\leq n}\delta_{2n}^2(X_{2i})\sum_{i=1}^{n}\sum_{j\neq i}^{n}\left|\overline{K}_H(X_i - X_j)\right| \\
&= O_p\left(n|H|^{1/2}\eta_1^2\eta_2^2\right).
\end{aligned}$$

*Lemma A.3*

$$C_{3n} \equiv n|H|^{1/2}\left\{\int A_{3n}^2(x)dx - B_{3n}\right\} = o_p(1).$$

*Proof.* Note that $n|H|^{1/2}B_{3n} = n^{-1}|H|^{1/2}\sum_{i=1}^{n}U_{1i}^2((m_2(X_{2i}) - \tilde{m}_2(X_{2i})))^2\overline{K}_H(0) = O_p(|H|^{-1/2}n^{-1}|H_2|^{-1}) = o_p(1)$. It suffices to show that $n|H|^{1/2}\int A_{3n}^2(x)dx = o_p(1)$, which holds provided that $A_{3n}(x) = o_p(n^{-1/2}|H|^{-1/4})$ uniformly in $x$. For later use, we show a stronger result, i.e., uniformly in $x$,

$$\begin{aligned}
A_{3n}(x) &= \frac{1}{n}\sum_{i=1}^{n}U_{1i}(\tilde{m}_2(X_{2i}) - m_2(X_{2i}))K_H(x - X_i) \\
&= o_p\left((n\log n)^{-1/2}\right).
\end{aligned} \tag{A.2}$$

Note that

$$\begin{aligned}
A_{3n}(x) &= \mathcal{V}_n(x) + \mathcal{B}_n(x) + O_p \\
&\qquad \left((\|H_2\| + n^{-1/2}|H_2|^{-1/2}\sqrt{\log n})\eta_2\right),
\end{aligned} \tag{A.3}$$

where

$$\mathcal{V}_n(x) = n^{-2}\sum_{i=1}^{n}\sum_{j=1}^{n}U_{1i}U_{2j}f_2^{-1}(X_{2i})K_{H_2}(X_{2i} - X_{2j})K_H(x - X_i),$$

$$\begin{aligned}
\mathcal{B}_n(x) = n^{-2}\sum_{i=1}^{n}\sum_{j=1}^{n}U_{1i}f_2^{-1}(X_{2i})b_2(X_{2j}, X_{2i})K_{H_2}(X_{2i} - X_{2j}) \\
K_H(x - X_i).
\end{aligned}$$

For fixed $x$, $\mathcal{V}_n(x)$ is a second-order $V$ statistic, and it is easy to show that $\mathcal{V}_n(x) = O_p(n^{-1}|H_2|^{-1/2}|H|^{-1/2})$. For a uniform bound on $\mathcal{V}_n(x)$, we can modify the proof of (A.10) in Gozalo and Linton (2001) to show that

$$\begin{aligned}
\sup_x|\mathcal{V}_n(x)| &= O_p\left(n^{-1}|H_2|^{-1/2}|H|^{-1/2}\log n\right) \\
&= o_p\left((n\log n)^{-1/2}\right).
\end{aligned} \tag{A.4}$$

where the last equality follows because $n|H_2||H|/(\log n)^2 \to \infty$ as $n \to \infty$. Now uniformly in $x$,

$$\begin{aligned}
\mathcal{B}_n(x) &= n^{-1}\sum_{i=1}^{n}U_{1i}K_H(x - X_i)f_2^{-1}(X_{2i})n^{-1} \\
&\quad \times\sum_{j=1}^{n}b_2(X_{2j}, X_{2i})K_{H_2}(X_{2i} - X_{2j}) \\
&= ((p_2 + 1)!n)^{-1}\sum_{i=1}^{n}U_{1i}K_H(x - X_i) \\
&\quad \times\left\{\sum_{j=1}^{d_2}h_{2j}^{p_2+1}\frac{\partial^{p_2+1}m(X_{2i})}{\partial X_{2i,j}}\gamma_k^{d_2} + O_p\left(\|H_2\|^{p_2+1+c_2}\right)\right\} \\
&= O_p\left(n^{-1/2}|H|^{-1/2}\sqrt{\log n}\|H_2\|^{p_2+1} + \|H_2\|^{p_2+1+c_2}\right) \\
&= o_p\left((n\log n)^{-1/2}\right),
\end{aligned} \tag{A.5}$$

where $\gamma_k = \int_{\mathbb{R}} u^{p_j+1}k(u)du$ and $c_2 = 1$ if $p_2$ is even and 2 if $p_2$ is odd. Equation (A.2) follows from (A.3)–(A.5) and the fact that $(\|H_2\| + n^{-1/2}|H_2|^{-1/2}\sqrt{\log n})\eta_2 = o((n\log n)^{-1/2})$. ■

*Lemma A.4*

$$C_{4n} \equiv n|H|^{1/2}\left\{\int A_{4n}^2(x)dx - B_{4n}\right\} = o_p(1).$$

*Proof.* The proof is analogous to that of Lemma A.3. ■

*Lemma A.5*

$$C_{5n} \equiv n|H|^{1/2}\int A_{1n}(x)A_{2n}(x)dx = o_p(1).$$

*Proof.* First,

$$\begin{aligned}
\sup_x|A_{2n}(x)| &= \sup_x\left|\frac{1}{n}\sum_{i=1}^{n}(m_1(X_{1i}) - \tilde{m}_1(X_{1i}))(m_2(X_{2i})\right. \\
&\qquad \left. - \tilde{m}_2(X_{2i}))K_H(x - X_i)\right| \\
&\leq C\eta_1\eta_2\sup_x\left|\frac{1}{n}\sum_{i=1}^{n}K_H(x - X_i)\right| = O_p(\eta_1\eta_2).
\end{aligned}$$

By the standard argument (e.g., Masry, 1996, Theorem 2),

$$\begin{aligned}
A_{1n}(x) &= \frac{1}{n}\sum_{i=1}^{n}U_{1i}U_{2i}K_H(x - X_i) \\
&= O_p(n^{-1/2}|H|^{-1/2}\sqrt{\log n}) \text{ uniformly in } x.
\end{aligned} \tag{A.6}$$

Consequently, $C_{5n} = O_p(n|H|^{1/2}\eta_1\eta_2 n^{-1/2}|H|^{-1/2}\sqrt{\log n}) = O_p(n^{1/2}\eta_1\eta_2\sqrt{\log n}) = o_p(1)$. ■

*Lemma A.6*

$$C_{6n} \equiv n|H|^{1/2}\int A_{1n}(x)A_{3n}(x)dx = o_p(1).$$

*Proof.* By (A.2) and (A.6), $C_{6n} = O_p(n|H|^{1/2} (n \log n)^{-1/2} n^{-1/2} |H|^{-1/2} \sqrt{\log n}) = o_p(1)$. ■

*Lemma A.7*

$$C_{7n} \equiv n|H|^{1/2} \int A_{1n}(x) A_{4n}(x) dx = o_p(1).$$

*Proof.* The proof is analogous to that of Lemma A.6. ■

*Lemma A. 8*

$$C_{8n} \equiv n|H|^{1/2} \int A_{2n}(x) A_{3n}(x) dx = o_p(1).$$

*Proof.* Note that $n|H|^{1/2} B_{2n} = O_p(|H|^{-1/2} \eta_1^2 \eta_2^2) = o_p(1)$, and $n|H|^{1/2} B_{3n} = O_p(|H|^{-1/2} n^{-1} |H_2|^{-1}) = o_p(1)$. It follows from Lemmas A.2 and A.3 that $n|H|^{1/2} \int A_{jn}^2(x) dx = o_p(1)$ for $j = 2, 3$. The result follows from the Cauchy-Schwartz inequality. ■

*Lemma A.9*

$$C_{9n} \equiv n|H|^{1/2} \int A_{2n}(x) A_{4n}(x) dx = o_p(1).$$

*Proof.* The proof is analogous to that of Lemma A.8. ■

*Lemma A. 10*

$$C_{10n} \equiv n|H|^{1/2} \int A_{3n}(x) A_{4n}(x) dx = o_p(1).$$

*Proof.* The proof is analogous to that of Lemma A.8. ■

## ACKNOWLEDGMENTS

The authors gratefully thank Serena Ng, Hashem Pesaran, the associate editor, and two anonymous referees for their many insightful comments that have led to great improvement of the presentation. The first author gratefully acknowledges financial support from the NSFC, under Grant Nos. 70501001 and 70601001, and from the Guanghua-Wharton Joint Research Fund. The second author gratefully acknowledges the financial support from the Academic Senate, UCR.

## REFERENCES

Alexander, C. (2001), "Orthogonal GARCH," in *Mastering Risk*. London: Prentice Hall, pp. 21–38.

Anatolyev, S. (2002), "Conditional and Unconditional Correlatedness and Heteroskedasticity," *Econometric Theory,* 18, 820–821.

Anselin, L. (1988). *Spatial Econometrics: Methods and Models*, Dordrecht: Kluwer Academic Publishers.

Bosq, D. (1998), *Nonparametric Statistics for Stochastic Processes: Estimation and Prediction*, Springer: New York.

Breusch, T. S., and Pagan, A. R. (1980), "The Lagrange Multiplier Test and its Applications to Model Specification Tests in Econometrics," *The Review of Economic Studies,* 47, 239–253.

Dufour, J.-M., and Khalaf, L. (2002), "Exact Tests for Contemporaneous Correlation of Disturbances in Seemingly Unrelated Regressions," *Journal of Econometrics,* 106, 143–170.

Duncan, G. M. (1983), "Estimation and Inference for Heteroscedastic Systems of Equations," *International Economic Review,* 24, 559–566.

Engle, R. F. (2002), "Dynamic Conditional Correlation—a Simple Class of Multivariate GARCH Models," *Journal of Business & Economic Statistics,* 20, 339–350.

Engle, R. F., and Kroner, K. F. (1995), "Multivariate Simultaneous Generalised ARCH," *Econometric Theory,* 11, 122–150.

Engle, R. F., and Sheppard, K. (2002), "Theoretical and Empirical Properties of Dynamic Conditional Correlation Multivariate GARCH," mimeo, University of San Diego, Dept. of Economics.

Fan, J., and Gijbels, I. (1996), *Local Polynomial Modelling and Its Applications*, New York: Chapman and Hall.

Fan, J., Wang, M., and Yao, Q. (2008), "Modelling Multivariate Volatilities via Conditionally Uncorrelated Components," *Journal of the Royal Statistical Society: Series B,* 70, 679–702.

Frees, W. W. (1995), "Assessing Cross-sectional Correlation in Panel Data," *Journal of Econometrics,* 69, 393–414.

Gao, J., and King, M. (2003), "Estimation and Model Specification Testing in Nonparametric and Semiparametric Regression Models," *mimeo*, University of Western Australia, Dept. of Mathematics and Statistics.

Gozalo, P., and Linton, O. (2001), "Testing Additivity in Generalized Nonparametric Regression Models with Estimated Parameters," *Journal of Econometrics,* 104, 1–48.

Hafner, C. M., and Herwartz, H. (2000), "Testing for Linear Autoregressive Dynamics under Heteroskedasticity," *The Econometrics Journal,* 3, 177–197.

Hall, P., and Carroll, R. J. (1989), "Variance Function Estimation in Regression: the Effect of Estimating the Mean," *Journal of the Royal Statistical Society B,* 51, 3–14.

Hamilton, J. D. (1994). *Time Series Analysis*. Princeton, NJ: Princeton University Press.

Härdle, W., Tsybakov, A., and Yang, L. (1998), "Nonparametric Vector Autoregression," *Journal of Statistical Planning and Inference,* 68, 221–245.

Horowitz, J. L., and Spokoiny, V. G. (2001), "An Adaptive, Rate-optimal Test of a Parametric Mean-regression Model against a Nonparametric Alternative," *Econometrica,* 69, 599–631.

——— (2002), "An Adaptive, Rate-optimal Test of Linearity for Median Regression Models," *Journal of the American Statistical Association,* 97, 822–835.

Kariya, T. (1981), "Tests for the Independence between Two Seemingly Unrelated Regression Equations," *Annals of Statistics,* 9, 381–390.

Kreiss, J. P. (1997), "Asymptotic Properties of Residual Bootstrap for Autoregessions," *manuscript*, Technical University of Braunschweig, Institute for Mathematical Statistics.

Lavergne, P., and Vuong, Q. (2000), "Nonparametric Significance Testing," *Econometric Theory,* 16, 576–601.

Lee, A. J. (1990). "*U-statistics: Theory and Practice*", New York: Marcel Dekker.

Lee, S. (2003), "Efficient Semiparametric Estimation of Partially Linear Quantile Regression Model," *Econometric Theory,* 19, 1–31.

Li, Q., and Hsiao, C. (2001), "A Consistent Test for Conditional Heteroskedasticity in Time-Series Regression Models," *Econometric Theory,* 17, 188–221.

Liu, R. Y. (1988), "Bootstrap Procedure under Some Non-iid Models," *Annals of Statistics,* 16, 1696–1708.

Mandy, D., and Martins-Filho, C. (1993), "Seemingly Unrelated Regressions under Additive Heteroskedasticity: Theory and Share Equation Applications," *Journal of Econometrics,* 58, 315–346.

Masry, E. (1996), "Multivariate Local Polynomial Regression for Time Series: Uniform Strong Consistency Rates," *Journal of Time Series Analysis,* 17, 571–599.

Moran, P. A. P. (1948), "The Interpretation of Statistical Maps," *Biometrika,* 35, 255–260.

Ng, S. (2006), "Testing Cross-section Correlation in Panel Data Using Spacings," *Journal of Business & Economic Statistics,* 24, 12–23.

Pesaran, M. H. (2004), "General Diagnostic Tests for Cross Section Dependence in Panels," IZA discussion paper no. 1240, University of Cambridge.

——— (2006), "Estimation and Inference in Large Heterogeneous Panels with Multifactor Error Structure," *Econometrica,* 74, 967–1012.

Pesaran, M. H., Ullah, A., and Yamagata, T. (2006), "A Bias Adjusted LM Test of Error Cross Section Independence," *working paper*, UCR.

Racine, J., and Li, Q. (2004), "Nonparametric Estimation of Regression Functions with Both Categorical and Continuous Data," *Journal of Econometrics,* 119, 99–130.

Robinson, P. M. (1987), "Asymptotically Efficient Estimation in the Presence of Heteroskedasticity of Unknown Form," *Econometrica,* 55, 875–891.

Ruckstuhl, A. F., Welsh, A. H., and Carroll, R. J. (2000), "Nonparametric Function Estimation of the Relationship between Two Repeatedly Measured Variables," *Statistica Sinica,* 10, 51–71.

Singh, B., and Ullah, A. (1974), "Estimation of Seemingly Unrelated Regressions with Random Coefficients," *Journal of the American Statistical Association,* 69, 191–195.

Smith, M., and Kohn, R. (2000), "Nonparametric Seemingly Unrelated Regression," *Journal of Econometrics,* 98, 257–281.

Srivastava, V., and Giles, D. (1987). *Seemingly Unrelated Regression Equations Models*, New York. Marcel Dekker.

Su, L., and White, H. (2007), "A Consistent Characteristic Function-based Test for Conditional Independence," *Journal of Econometrics,* 141, 807–834.

Welsh, A. H., and Yee, T. W. (2006), "Local Regression for Vector Responses," *Journal of Statistical Planning and Inference,* 136, 3007–3031.

Wu, C. F. J. (1986), "Jackknife, Bootstrap and Other Resampling Methods in Regression Analysis," *Annals of Statistics,* 14, 1261–1295.

Xu, K. L. (2006), "Bootstrapping Autoregression under Nonstationary Volatility," *mimeo*, Yale University, Dept. of Mathematics.

Zellner, A. (1962), "An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias," *Journal of the American Statistical Association,* 57, 348–368.